What is the SAC role in identifying and resolving data quality issues found in secondary data sets?

Presented by
Justice Research and Statistics Association

January 28, 2021
Hate & Bias Crime Research in Maine: Data Quality Controls

George Shaler
Hannah Brintlinger
Jack McDevitt
Project Goal

To establish clearance rates for reported hate and bias crimes from 2008 to 2017 in Maine.
Hate & Bias Crimes in Maine

- From 2008 to 2012, Maine averaged 58.6 bias/hate crime incidents annually across the state. This average fell to 33.2 for the 2013 to 2017 period, a 43.3% drop.

- Nationally, the average number of incidents fell as well during this same period; however the decline was much smaller at 8.7%. (National rates obtained from https://ucr.fbi.gov/hate-crime)
Challenges with Hate & Bias Crime Research

- Incidents are frequently not reported to the police.
- Law enforcement agencies must recognize any indications of bias, document their findings as bias crimes, and then report the incident as a bias crime to the UCR.
- Very few states’ hate crime reports contain information on arrests or convictions.
Hate Crimes Justice Process in Maine

- State Attorney General can bring a civil order against the accused.
- District Attorneys can bring criminal proceeding against the accused for other offenses committed during the perpetration of the crime.
Methods

- Compile a list of incidents that occurred in Maine from FBI Crime Explorer Dataset

- Request information from all Maine LE agencies that reported incidents to the UCR during the study period
  - Did they have record of the incident?
  - Did an arrest occur?
  - Arrest Tracking Number
  - Town the incident occurred in (more for county sheriffs)
  - Was a weapon used?
  - Was medical care required?

- Work with the DAs and the State AG to figure out case outcomes for those incidents that resulted in arrest
Methods - Continued

- 445 incidents were generated by downloading these data from the FBI’s Crime Explorer.

- Incident level data were sent to each of the 65 agencies (42% of all agencies in ME) that reported a hate crime during the study period.
Methods - Challenges

- Some of the records chosen are located in old record management systems.
- Some records in FBI’s Crime Explorer could not be located or were not known to local law enforcement.
- Getting data from local law enforcement, the District Attorneys, and the AG’s Office.
Methods – Challenges (2)

- Large number of data sets
  - 62 different sources.
    - Initial dataset - FBI data
    - Data from 51 local law enforcement agencies
    - State AG data
    - 8 District Attorneys
    - Department of Public Safety for arrest tracking numbers (ATNs) and Incident #s.
Data Collection

- Sent 65 data requests to LE agencies across Maine
  - 11 agencies never got back to us
  - 3 agencies told us they could not locate any of the incidents
  - 51 (78.5%) agencies provided us with information
Data Quality Issues

- Sent data requests out for 445 incidents
  - Note: some agencies provided us with information on incidents that were not in the FBI dataset. We decided to include those incidents in the dataset moving forwards.
- We received LE incident information for 399 (89%) incidents
Data Quality Issues – Continued

- The State AG’s Office is providing the Maine SAC with updates on any of the incidents that resulted in an arrest and whether the AG’s Office reached a civil order with defendant.

- The 8 ME DAs are providing the Maine SAC with information on those arrested and whether a conviction was reached on non-hate charges.

- Both of these agencies need an ATN to search their databases for follow-up information. Sometimes the ATNs supplied by law enforcement are not accurate.
Next Steps

- Follow-up with 8 DAs regarding the arrests
  - How many were prosecuted
  - # of convictions
- Follow-up with State AG
  - # of civil orders
  - Other actions
- Focus groups with law enforcement agencies regarding trainings
Victim-Offender Relationship Coding in Virginia Incident-Based Uniform Crime Reports 2006-2015

Presented by
Jim McDonough, Ph.D.
Manager, Criminal Justice Research Center
Virginia Department of Criminal Justice Services
January 2021

This project was supported by Grant No. 2016-BJ-CX-K005 awarded by the Bureau of Justice Statistics, Office of Justice Programs, U.S. Dept. of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Dept. of Justice.
SAC - VA State Police Relationship

- SAC has working relationship with Virginia’s UCR Repository since VA adopted statewide VAIBRS

- SAC is located in the VA Dept. of Criminal Justice Services, VA’s State Administering Agency (SAA)

- VAIBRS Repository is located in the Virginia State Police Uniform Crime Reporting section

- The SAC receives monthly electronic copies of the VAIBRS files produced by State Police

- SAC uses VAIBRS data to help guide many of SAA’s policy, planning and grant-making activities
Virginia Incident-Based Crime Reporting System (VAIBRS)

- Virginia law enforcement agencies began reporting incident-based uniform crime data to the Virginia State Police repository in late 1990s

- 100% IBR coverage since 2005

- Participating agencies including county sheriffs, city and county police departments, state police, college/university police, “other” agencies

- VAIBRS follows same reporting rules/codes as NIBRS
SAC VAIBRS Data Quality/Completeness Checks

- SAC VAIBRS analyst reviews each monthly file
- Examines standard frequency distributions for outliers
- Compares numbers/types of annual reports to reports from prior years to look for drops in reporting
- Looks at distributions for state level aggregate data and for individual agencies
- Reviews are not “audit samples”
- Final check done on annual “frozen” file with all updates, deletions, etc.
Virginia SAC examined victim-offender relationships (VORs) reported in VAIBRS from 2006 through 2015

- Assess completeness and consistency of VORs
  - for all victims of person offenses or robbery, and for victims of serious violence (murder, rape, robbery, aggravated assault)

- Determine distribution of known VORs (Intimate, Family, Acquaintance-Friend, Stranger) vs. unknown VORs (“RU” and missing or null)

- Identify changes in VOR reporting over time
Study Description

- Analyzed VAIBRS reports from 120 agencies (96% of all victims reported)

- 1.1 million victims of person offense or robbery from 2006 through 2015
  - 16% (N=187,270) victims of serious violence (murder, rape, robbery, aggravated assault)

- 23% of victims associated with multiple offenders and have multiple VORs coded
  - "Primary" VOR selected to examine—highest level of emotional intensity (Decker 1993)
Distribution of VORs in 2005-2016

Number of Victims

Percent of Known Victims VOR
Unknown VORs: “RU” or missing

Some level of unknown VORs is expected. Typical scenarios explaining unknown VORs:

- No suspect when initial incident report submitted to VAIBRS so no associated offender record; record submitted with VOR missing or null

- Suspect is known and offender record submitted but relationship is unknown = VOR coded “RU”

- VOR known but agency chooses not to report it
Unknown VORs: “RU” or missing

Victim demographic likeliest to have an unknown VOR:

- Male
- “Other” race
- Hispanic
- Ages 13-17 or 65 and older
Unknown VORs: “RU” or missing, by Offense Type

Victims of robbery and homicide likeliest to have an unknown VOR

<table>
<thead>
<tr>
<th>Offense</th>
<th>% VOR Unknown</th>
<th>% VOR Known</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robbery N=67,869</td>
<td>54.0</td>
<td>46.0</td>
</tr>
<tr>
<td>Murder N=3,380</td>
<td>48.4</td>
<td>51.6</td>
</tr>
<tr>
<td>Other Homicide N=276</td>
<td>37.0</td>
<td>63.0</td>
</tr>
<tr>
<td>Aggravated Assault N=93,631</td>
<td>30.3</td>
<td>69.7</td>
</tr>
<tr>
<td>Intimidation N=113,116</td>
<td>26.1</td>
<td>73.9</td>
</tr>
<tr>
<td>Kidnapping N=7,030</td>
<td>22.7</td>
<td>77.3</td>
</tr>
<tr>
<td>Fondling N=24,869</td>
<td>22.2</td>
<td>77.8</td>
</tr>
<tr>
<td>Rape N=22,390</td>
<td>20.4</td>
<td>79.6</td>
</tr>
<tr>
<td>Total N=1,141,646</td>
<td>19.0</td>
<td>81.0</td>
</tr>
<tr>
<td>Statutory Rape N=1,652</td>
<td>14.4</td>
<td>85.6</td>
</tr>
<tr>
<td>Simple Assault N=807,448</td>
<td>13.5</td>
<td>86.5</td>
</tr>
<tr>
<td>Incest N=250</td>
<td>10.8</td>
<td>89.2</td>
</tr>
</tbody>
</table>
VOR Consistency and Completeness

- How well VORs coded in accordance with reporting rules, for example:
  - In NIBRS, Spouse (VOR="SE") victim (and offender) must be at least 14 years old
  - Common sense rules: Parent victim (VOR="PA") age greater than child offender (VOR="CH") age (of course, rare exceptions)

- How complete were associated victim and offender elements reported
  - Inconsistent elements may indicate miscoded VOR
### VOR Consistency and Completeness

<table>
<thead>
<tr>
<th>Element</th>
<th>% Consistent</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>VOR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BG, CS, HR, PA, SE, or XS</td>
<td>98-99</td>
<td>Boyfriend/girlfriend, common law spouse, homosexual relationship, parent, spouse, ex-spouse</td>
</tr>
<tr>
<td>CH, GC, or GP</td>
<td>95-97</td>
<td>Child, grandchild, grandparent</td>
</tr>
<tr>
<td>SP (Stepparent)</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>CF (child of boyfriend/girlfriend)</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>Victim age/sex/race complete</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>Offender age/sex/race complete</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>Incest victims have family VOR</td>
<td>86</td>
<td></td>
</tr>
</tbody>
</table>

- Almost all (97-99%) VORs reported were consistent with associated victim/offender elements (age, sex) for:
  - Intimate partner
  - Family (parent, child, grandchild, grandparent)

- Consistency issues found with VOR “CF” (child of boyfriend/girlfriend), “SP” (stepparent), and VORs for incest victims
VOR Consistency and Completeness Issues

VOR “SP” (victim stepparent)

- Age of victim and offender suggests:
  - 15% offender-victim relationship was coded instead of victim-offender relationship; victim was stepchild (VOR=“SC”) (VOR reversed)
  - 14% victim may be spouse (VOR=“SE”)

VOR “CF” (victim child of boyfriend/girlfriend)

- 60% of victims too close in age to offender
  - Some portion may be miscoded “Other family” relationships (VOR=“OF”)

Nine of 250 incest victims with “RU” VOR code
# Agency-level VOR Reporting

120 agencies stratified into nine groups adapted from FBI agency population groups

<table>
<thead>
<tr>
<th>FBI Population Group</th>
<th>Study Population Group</th>
<th>Number of Agencies (N=120)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1--Cities 250,000 and over</td>
<td>n/a</td>
<td>0</td>
</tr>
<tr>
<td>Group 2--Cities 100,000 to 249,999</td>
<td>Cities 100,000 to 499,999</td>
<td>7</td>
</tr>
<tr>
<td>Group 3—Cities 50,000 to 99,999</td>
<td>Cities and towns 50,000 to 99,999</td>
<td>6 (5 cities, 1 town)</td>
</tr>
<tr>
<td>Group 4--Cities 25,000 to 49,999</td>
<td>Cities and towns 25,000 to 49,999</td>
<td>8 (7 cities, 1 town)</td>
</tr>
<tr>
<td>Group 5--Cities 10,000 to 24,999</td>
<td>Cities and towns 10,000 to 24,999</td>
<td>16 (12 cities, 4 towns)</td>
</tr>
<tr>
<td>Group 6--Cities 2,500 to 9,999</td>
<td>Cities and towns 2,500 to 9,999</td>
<td>14 (4 cities, 10 towns)</td>
</tr>
<tr>
<td>Group 7--City &lt;2,500 and college-university, and “other” agencies with no population</td>
<td>Cities and towns &lt; 2,500 and college, university, and “other” agencies with no population</td>
<td>1 university</td>
</tr>
<tr>
<td>Group 8--Non-metro County (includes State Police)</td>
<td>Same as FBI group</td>
<td>26</td>
</tr>
<tr>
<td>Group 9--Metro County (includes State Police)</td>
<td>Same as FBI group</td>
<td>42</td>
</tr>
</tbody>
</table>

Compare agency reporting across state and within population group: proportion VOR known
Percent VOR Known by Agency
All person offenses, 2006-2015

N= 120 agencies

99 (82%) reported from 80% to 99% known primary VORs

3 (2%) reported more than half of all VORs as unknown or null
Percent Known VOR by Agency Group

Group 2
Cities population 100K+

Group 3
Cities 50K-99K

Group 4
Cities 25K-49K

Group 5
Cities 10K-24K

Groups 6 & 7
Cities 2.5K-9K and an urban university

Group 8
Non-metro counties

Group 9
Metro counties
One Agency Outlier Reported more than 50% of VORs as “RU”

Group 9 Agency09-G9
A police department serving one of the largest metropolitan counties in Virginia

- This agency’s acquaintance, stranger, “RU” and missing VOR reporting patterns differed significantly from all others in Group 9
- Highest proportion (and number) of unknown VORs overall
- SAC reported this anomaly to UCR repository analyst
VSP UCR Repository Contacted Reporting Agency with 50%+ VORs coded “RU”

Agency09_G9 response to large number of unknown VORs:

- Department’s records management system requires a VOR for “domestic assaults”
- Codes “RU” for relationships without available codes such as “child in common but are not married or boyfriend/girlfriend” or “ex boyfriend-girlfriend”
- When VOR is not entered, system defaults to “RU” (accounts for no missing or null VORs reported)
- Agency did not address the unusually low number of “Acquaintance” and “Stranger” VORs
Summary

- SAC and VSP UCR section have good working relationship – analysts in both shops communicate with one another

- VSP keeps SAC apprised of changes in data elements, coding values, instructions to local reporting agencies – anything that would affect SAC’s interpretation of the data

- SAC notifies VSP whenever its data analyses indicate a possible issue with data quality or completeness

- SAC’s detailed data analyses sometimes spot issues missed by VSP’s standard error checks
Summary

- Whenever SAC identifies a data issue pertaining to a single reporting locality, we notify VSP, not the locality, to look into it.

- SAC respects VSP’s data collection role and need to maintain a good working relationship with the local reporting law enforcement agencies.

- SAC does not want to get between VSP and the local agencies that report to VSP.

- VSP sometimes comes to SAC when it notices some irregularity in the data, but needs a detailed analytic look.
Datasets and Data

Illinois Criminal Justice Information Authority (ICJIA)

Justin Escamilla, PhD
Acting Manager
Center for Criminal Justice Data and Analytics
Research and Analysis Unit
Overview

The Data Warehouse

Datasets

Examples of Data Quality Issues

Data Quality Checking

Quality Checkpoints

What’s to come?
The Data Warehouse

Who and what
• Four full-time staff
• Yearly grant project (SJS)
• Publish criminal justice datasets to web
• Respond to data requests

Recent shifts
• New processing and formats for datasets published to web
• New data updating software
• Staff changes
• Access to criminal history records updated
Datasets

Research Hub

• 24 datasets on corrections, courts, crime, law enforcement, victims, etc…

• Prison admission and exit, parole admissions
  • Source: Illinois Department of Corrections

• Criminal court case filings
  • Source: Administrative Office of the Courts website

• Uniform Crime Report (UCR)
  • Source: State police website

Internal

• Criminal History Record Information data
  • Source: Restricted access to state police data
    • SQL server
Examples of Data Quality Issues

- Uniform Crime Report
  - Yearly data updated with new “Crime in Illinois” reports, which includes revised crime numbers for year prior

- Juvenile court case filing
  - Recent years correctly matched PDF reports, but the years prior were incorrectly flipped

- CHRI pull
  - After data system transfer, weekly totals of arrests declined to zero; expungements

- Jail detention data
  - Numbers reported by us did not match reports from other sources
Data Quality Checking

- Web dataset checking not formalized
  - Large errors easily noticed will trigger review
    - Not systematically done
  - Users notifying us of an issue will trigger review
- CHRI checking is more regular and sophisticated
Data Checking Questions (CHRI)

- How did the total number of records change from last year to this year?
- How do totals from an example pull change from last year to this year?
- How do totals from an example pull in the annual pulls compare to the live system?
- Does each arrest have at least one arrest charge?
- Does each record from state’s attorney, court disposition, charge, or sentence data have an arrest?
- Does each sentence record link to a court disposition?
- Do court dispositions and arrests show the same name?
- Are the characteristics recorded in arrests consistent?
- Felony conviction connection with a new prison admission
- Does name/DOB combo consistently match with more than one State ID?
Quality Checkpoints

- Source data
  - Missing data/poor reporting
  - Inaccurate or unexplained collection
  - Confusing dissemination
- Pulldown of source data
  - Corrupted or mislabeled files
  - Errors in data entry, saving or reformatting
Quality Checkpoints (cont.)

- Recoding or aggregation activities
  - Coding errors
- Publication/reporting
  - Mislabeling
  - Mischaracterizations
Future

- Data publishing/sharing checklists covering
  - Data sources
  - Checkpoints
- CHRI audit revamp
- Arrest explorer

Step 1
1. 
2. 
3.

Step 2
1. 
2. 
3.

Step 3
1. 
2. 
3.
Acknowledgements

- Ernst Melchior
- Jessica Reichert
- Christine Devitt Westley